

KALMAN FILTERS FOR PROCESSES WITH
UNKNOWN INITIALVALUES

PER HAGANDER

Report 7332 September 1973
Lund Institute of Technology
Division of Automatic Control



KALMAN FILTERS FOR PROCESSES WITH UNKNOWN INITIAL VALUES.

P. Hagander

ABSTRACT.

A Kalman filter needs an à priori statistics for the initial state. It is shown how the filter should be started if some part of the initial state is totally unknown. The duality with optimal control with end point constraints is very useful both for proofs and intuition.

The usual way of starting with a very large covariance has very bad numerical properties. The optimal discrete time filter is determined by two "Riccati equations", one matrix to keep track of the bias until the unknown initial value is observable, and one matrix for the error covariance.

In continuous time the estimation is more complicated. The whole system becomes observable at once. After an initial discontinuity a usual Kalman filter could be started, but the gain would be almost infinite. It is therefore suggested how the estimate should be calculated using a separation into two estimates. The optimal linear stochastic control is also discussed.



TABLE OF CONTENTS

1. Introduction
2. Problem formulation and main theorem
3. Separation into two estimates
4. Measurement noise only
5. Combination of two filters
6. Linear stochastic control
7. Continuous time, duality
8. Conclusions

Appendix

References

1. INTRODUCTION.

State estimation for linear stochastic systems is a well established theory, see e.g. [2, 5]. The Kalman filter requires, however, a known statistics for the initial state, which is often not available. The suboptimal solution recommended for instance by Sorensen [9] seems to be the most accepted substitute. The covariance of the initial state is assumed to be a unit matrix times a scalar, which is large compared with all other covariances. The resulting estimate and especially the assumed error covariance may, however, deviate considerably from the correct values. A serious fact is also that the Riccati equation (discrete or continuous time) will be ill conditioned if the initial covariance is very large.

Here it will be shown how to obtain the best linear unbiased estimate, i.e. the minimal variance estimate under the constraint that the expected value of the error should be zero for all initial states. Of course, it is only possible to obtain such an estimate for observable systems.

Such estimates were discussed by Kalman [5] for the case with only measurement noise, but they do not seem to have received much attention since, probably because the algebra is repellent. In the context of least squares parameter estimation, there has appeared some related work, e.g. [1].

In the following section the problem will be defined rigorously, and the solution for discrete time is obtained by letting the initial covariance reach infinity. The bad numerical properties of large covariances will be obvious, and the necessity of two Riccati equations is demonstrated.

In order to give a better understanding of the estimate and the necessary matrices, another derivation is made using a se-



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41545321_0010

paration into two estimates. In Section 3 the separation is shown, giving one filter for the stochastic terms and one for the unknown initial terms that has only measurement noise. The latter case is treated in Section 4, where recursive equations are given also before the system has become observable. It is shown in Section 5 how the two filters combine to one, the same as obtained in Section 2.

The filter is a time variable "dead beat" filter, which is especially simple in the single output case. Two matrix recursions are needed to get the filter gain. One matrix is the error covariance, and the other spans the bias of the estimate. As soon as the unknown initial value is observable the bias is zero, and the filter will continue as the usual Kalman filter.

In Section 6 the separation principle is applied, and the linear stochastic regulator problem is discussed.

Finally, it is shown that the problem is the dual of optimal control for fixed end state, and the solution for continuous time is obtained in this way. The difficulties with the differential equation formulation are discussed for time invariant systems, and a smoothing type algorithm is recommended for continuous time.

2. PROBLEM FORMULATION AND MAIN THEOREM.

Consider the discrete time system

$$\begin{aligned} x(t+1) &= \phi x(t) + v(t) & x(t_0) &= x_0 \\ y(t) &= \theta x(t) + e(t) \end{aligned} \quad (2.1)$$

where v and e are uncorrelated white noise sequences with covariances R_1 and R_2 . For the initial state probability will be introduced in various degrees:

$$x_0 = x_0^S + x_0^N$$

The only assumption about x_0^N is that it is restricted to a subspace spanned by the full column rank rectangular matrix N^T :

$$x_0^N = N^T \xi, \quad \xi \text{ arbitrary}$$

x_0^S is uncorrelated with v and e and has zero mean value and covariance R_0^S . A very natural assumption, which will nowhere be used, is that x_0^S is restricted to a subspace disjoint from the range space of N^T .

Best linear unbiased estimate:

Introduce Y_t as the function $y(s)$, $s \in [t_0, \dots, t]$. It is now interesting to express Y_t as a linear function of ξ and α_t , a process into which all introduced random variables are collected. α_t has zero mean value and covariance Q_t .

$$Y_t = W_t \xi + \alpha_t \quad (2.2)$$

A linear unbiased estimate of ξ is a function F_t of Y_t such that

$$EF_t Y_t = \xi$$

for all values of ξ . The minimal variance unbiased estimator is given by the well-known Gauss-Markov Theorem, see e.g. [1, 7], provided that Q_t is nonsingular

$$\hat{\xi}(t) = (W_t^T Q_t^{-1} W_t)^{-1} W_t^T Q_t^{-1} Y_t \quad (2.3)$$

It should be noted that if ξ is assumed to be a random variable independent of all other introduced random variables, with covariance $\lambda^{-2}I$ and zero mean value, the Projection Theorem, e.g. [7], gives

$$\hat{\xi} = R_{\xi Y} R_Y^{-1} Y = \lambda^{-2} I W^T (W \lambda^{-2} I W^T + Q)^{-1} Y = (W^T Q^{-1} W + \lambda^2 I)^{-1} W^T Q^{-1} Y \quad (2.3a)$$

and (2.3) is obtained as the limit when $\lambda \rightarrow 0$, infinite initial covariance.

This demonstrates the equivalence between the minimal variance estimate for infinite a priori covariance and the minimal variance unbiased estimate.

If the system (2.1) has only white measurement noise, (2.3) directly solves the linear unbiased estimate of $x(t)$. Since $x(t)$ is a deterministic linear function of t , it follows that

$$\hat{x}(t|t) = \phi(t, t_0) N^T \hat{\xi}(t) \quad (2.4)$$

where

$$\phi(t, t_0) = \prod_{s=t_0}^{t-1} \phi(s+1, s)$$

Eq. (2.3) is also much simpler than in the general case because α and thus Q are especially simple. The inverse required in (2.3) is an observability Gramian.

In the general case (2.1) it is more complicated to obtain $\hat{x}(t|t)$. The system can be written as

$$\begin{cases} Y_t = W_t \xi + \alpha_t \\ x(t) = \phi(t, t_0) N^T \xi + \beta(t) \end{cases} \quad (2.5)$$

where α and β are correlated, and the state estimate will be

$$\hat{x}(t|t) = \phi(t, t_0) N^T \hat{\xi} + \hat{\beta}(t) \quad (2.6)$$

Operator formulas like in [4] could be used to evaluate (2.6), or the problem could be converted to a problem with only measurement noise, see Section 3. A third possibility is to use the limit argument of (2.3a) applied to the usual Kalman problem.

The infinite covariance limit of the Kalman filter.

If again x_0^N of (2.1) is a stochastic variable independent of x_0^S , v and e and with covariance $\lambda^{-2} N^T N$, a minimal variance unbiased estimate for the original problem could then be obtained by letting λ go to zero. The usual Kalman filter gives the minimal variance estimate:

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)[y(t) - \theta \hat{x}(t|t-1)] \quad (2.7)$$

$$\hat{x}(t+1|t) = \phi \hat{x}(t|t) \quad \hat{x}(t_0|t_0-1) = 0$$

$$K(t) = P(t) \theta^T (\theta P(t) \theta^T + R_2)^{-1}$$

$$P(t+1) = \phi [P(t) - K(t) \theta P(t)] \phi^T + R_1$$

$$P(t_0) = R_0 = R_0^S + \lambda^{-2} N^T N$$

Theorem 1: The minimal variance unbiased linear estimate for (2.1) is obtained by (2.7) and $K(t)$ from

$$K = \Lambda \theta^T (\theta \Lambda \theta^T)^+ [I - (\theta P_m \theta^T + R_2) [A(\theta P_m \theta^T + R_2) A]^+] + \\ + P_m \theta^T [A(\theta P_m \theta^T + R_2) A]^+ \quad (2.8)$$

with

$$A = I - (\theta \Lambda \theta^T)^+ (\theta \Lambda \theta^T) \quad (2.9)$$

$$\Lambda(t+1) = \phi [\Lambda - \Lambda \theta^T (\theta \Lambda \theta^T)^+ \theta \Lambda] \phi^T \quad \Lambda(t_0) = N^T N \quad (2.10)$$

$$P_m(t+1) = R_1 + \phi \left\{ (I - K \theta) P_m (I - K \theta)^T + K R_2 K^T \right\} \phi^T \\ P_m(t_0) = R_0^S \quad (2.11)$$

M^+ denotes the Moore Penrose pseudo inverse of M .

The estimate will be unbiased only if $\Lambda(t) = 0$.

Proof: Use induction in t to show that

$$P(t) = \lambda^{-2}\Lambda(t) + P_m(t)$$

which is true for $t = t_0$.

Introduce the full rank decompositions

$$\Theta\Lambda(t)\Theta^T = U^T U \quad \Theta P_m(t)\Theta^T = V^T V \quad V^T V + R_2 = H^T H$$

and consider the inverse

$$[\Theta P \Theta^T + R_2]^{-1} = [\lambda^{-2} U^T U + H^T H]^{-1}$$

which could be rewritten using a pseudo inverse formula given by Cline [3], see also [1].

$$= [(\bar{H}^T \bar{H})^+ + \lambda^2 (I - \bar{H}^+ H)(U^T U)^+ (I - \bar{H}^+ H)^T - \\ - \lambda^4 (I - \bar{H}^+ H)(U^T U)^+ H^T Q M(\lambda) Q H (U^T U)^+ (I - \bar{H}^+ H)^T]$$

with

$$A = I - U^+ U \quad \bar{H} = H A \quad Q = I - \bar{H} \bar{H}^+ \quad M(\lambda) = [I + \lambda^2 Q H (U^T U)^+ H^T Q]^{-1}$$

Note also that

$$U A = 0 \quad U [\bar{H}^T \bar{H}]^+ = 0 \quad U (I - \bar{H}^+ H) = U$$

so that

$$K = P \Theta^T [\Theta P \Theta^T + R_2]^{-1} = \Lambda \Theta^T (U^T U)^+ (I - \bar{H}^+ H)^T + P_m \Theta^T (\bar{H}^T \bar{H})^+ + \\ + \lambda^2 P_m \Theta^T (I - \bar{H}^+ H)(U^T U)^+ (I - \bar{H}^+ H)^T - \\ - \lambda^2 \Lambda \Theta^T (U^T U)^+ H^T Q M(\lambda) Q H (U^T U)^+ (I - \bar{H}^+ H)^T + O(\lambda^4)$$

and

$$K \Theta P = \lambda^{-2} \Lambda \Theta^T (U^T U)^+ \Theta \Lambda + \Lambda \Theta^T (U^T U)^+ (I - \bar{H}^+ H)^T \Theta P_m + \\ + P_m \Theta^T (\bar{H}^T \bar{H})^+ \Theta P_m + P_m \Theta^T (I - \bar{H}^+ H)(U^T U)^+ \Theta \Lambda - \\ - \Lambda \Theta^T (U^T U)^+ H^T Q M(\lambda) Q H (U^T U)^+ \Theta \Lambda + O(\lambda^2)$$

Thus for very small λ , $P(t+1)$ could be written

$$P(t+1) = \lambda^{-2}\Lambda(t+1) + P_m(t+1) \quad (2.12)$$

with $\Lambda(t+1)$ from (2.10)

$$\Lambda(t+1) = \phi \{ \Lambda - \Lambda \theta^T (\theta \Lambda \theta^T)^+ \theta \Lambda \} \phi^T$$

$$\begin{aligned} P_m(t+1) = & R_1 + \phi \{ P_m - \Lambda \theta^T (U^T U)^+ (I - \bar{H}^+ H)^T \theta P_m - \\ & - P_m \theta^T (I - \bar{H}^+ H) (U^T U)^+ \theta \Lambda - P_m \theta^T (\bar{H}^T \bar{H})^+ \theta P_m + \\ & + \Lambda \theta^T (U^T U)^+ H^T Q M(\lambda) Q H (U^T U)^+ \theta \Lambda \} \phi^T \end{aligned}$$

and

$$K(t) = \Lambda \theta^T (U^T U)^+ (I - \bar{H}^+ H)^T + P_m \theta^T (\bar{H}^T \bar{H})^+$$

which directly gives (2.8).

The induction is completed.

It is also clear from (2.12) that P will be large if Λ is not zero. In the limit this means that the system is not unbiased.

In order to prove (2.11) note that

$$(I - K\theta)P_m(I - K\theta)^T + K R_2 K^T = P_m - K\theta P_m - P_m \theta^T K^T + K(\theta P_m \theta^T + R_2)K^T$$

So it remains to show that

$$P_m \theta^T (\bar{H}^T \bar{H})^+ \theta P_m + \Lambda \theta^T (U^T U)^+ H^T Q M(0) Q H (U^T U)^+ \theta \Lambda = K H^T H K^T$$

But rewrite the right hand side using:

$$P_m \theta^T (\bar{H}^T \bar{H})^+ H^T H (\bar{H}^T \bar{H})^+ \theta P_m = P_m \theta^T (\bar{H}^T \bar{H})^+ \theta P_m$$

$$P_m \theta^T (\bar{H}^T \bar{H})^+ H^T H [I - (\bar{H}^T \bar{H})^+ H^T H] (U^T U)^+ \theta \Lambda = 0$$

$$\begin{aligned} \Lambda \theta^T (U^T U)^+ [I - H^T H (\bar{H}^T \bar{H})^+] H^T H [I - (\bar{H}^T \bar{H})^+ H^T H] (U^T U)^+ \theta \Lambda = \\ = \Lambda \theta^T (U^T U)^+ [H^T H - H^T H (\bar{H}^T \bar{H})^+ H^T H] (U^T U)^+ \theta \Lambda \end{aligned}$$

and the second term of the left hand side:

$$\begin{aligned} H^T Q M(0) Q H = H^T (I - \bar{H} \bar{H}^+) H = H^T H - H^T \bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T H = \\ = H^T H - H^T H (\bar{H}^T \bar{H})^+ H^T H \end{aligned}$$

which completes the proof of (2.11) and the whole theorem.

□

Example 2.1:

$$\phi = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \theta = [1 \quad 0 \quad 0] \quad R_1 = 0 \quad R_2 = 1$$

$$P(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda^{-2} \end{bmatrix} \quad K(0) = \begin{bmatrix} 1/2 \\ 0 \\ 0 \end{bmatrix}$$

$$P - K\theta P = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda^{-2} \end{bmatrix} \quad P(1) = \begin{bmatrix} 3/2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$K(1) = \begin{bmatrix} 3/5 \\ 2/5 \\ 0 \end{bmatrix} \quad P - K\theta P = \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.4 & 0.6 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$P(2) = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0.6 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$K(2) = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\} \cdot \frac{1}{3 + \lambda^{-2}} \rightarrow \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$P - K\theta P \approx (\lambda^{-2} - \lambda^{-2}) \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} [1 \quad 2 \quad 1] + \begin{bmatrix} 1 & 2 & 1 \\ 2 & 8.6 & 5 \\ 1 & 5 & 3 \end{bmatrix} \quad \lambda \rightarrow 0$$

which implies subtraction of very large numbers. This is avoided by theorem 1. The two terms of $P - K\theta P$ are stored separately in Λ and P_m .

In order to be able to use calculation by hand the example is very much simplified, and the illconditioness might

seem reasonable, but for a real system there is no significance left after a few such subtractions. The gain K will contain serious errors, and the real error covariance will not decrease although P does.

Example 2.2: θ is changed to

$$\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then

$$K(0) = \theta^T \cdot 1/2 \quad P - K\theta P = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & \lambda^{-2} \end{bmatrix}$$

$$P(1) = \frac{1}{2} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$K(1) = \left\{ \frac{1}{2} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} + \lambda^{-2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \right\} \left\{ \begin{bmatrix} 2 & 1/2 \\ 1/2 & 3/2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \lambda^{-2} \end{bmatrix} \right\}^{-1} \rightarrow$$

$$\rightarrow \frac{1}{4} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} \quad \lambda \rightarrow 0$$

after very illconditioned operations!

Comments: The new formula (2.8) may be seen as a way of improving the numerical condition of the calculations by keeping track of the large terms using the new matrix Λ . When Λ has become zero the filter is identical with the

usual Kalman filter, and the error covariance is $P = P_m$. Note that (2.11) is the form of the covariance updating formula that is valid for any K . It is not possible to rewrite (2.11) as a simple Riccati equation.

The interpretation of the optimal gain K in (2.8) is most obvious in the single output case. Then A is either 1 or 0, and

$$K = \begin{cases} \Lambda \theta^T (\theta \Lambda \theta^T)^{-1} & \text{if } A = 0 \\ P_m \theta^T (\theta P_m \theta^T + R_2)^{-1} & \text{if } A = 1 \end{cases} \quad (2.13)$$

If $A = 0$ the measurement contributes to the observability of ξ , and if $A = 1$ the information is only used to improve the current estimate just as in a Kalman filter.

Often the whole x_0 is unknown, i.e. N is a square matrix. Then $A = 0$ and $K = \Lambda \theta^T (\theta \Lambda \theta^T)^{-1}$ until $t = n$ and $\Lambda(n) = 0$, where n is the order of the system. All the first n measurements contribute to the observability. The filter could be called a dead-beat filter with time varying gain in analogy with dead-beat controllers. The influence of the unknown initial value on the estimation error at $t = n$ is zero. The estimate $\hat{x}(n|n-1)$ is unbiased. In this special case there are in fact no other unbiased estimates at $t = n$. Even the time constant dead-beat filter would have given the same $\hat{x}(n|n-1)$. It can be shown that the gain of the time constant filter is $K_c = K(n-1) = \Lambda(n-1) \theta^T (\theta \Lambda(n-1) \theta^T)^{-1}$. When the dead-beat filter is not unique, there is freedom left to minimize the error covariance. This is why the more complicated expressions (2.8) and (2.13) should be used in the multi-output case and when only part of the initial state is unknown. The filters are still time variable dead-beat filters!

The simple formulas (2.13) can be used also in the multi-output case, if the noise of the elements in the output vector are uncorrelated, i.e. if R_2 is diagonal. The elements can be used one at a time to update the estimate.

In order to give an interpretation of Λ and of the estimates for times before $\Lambda(t) = 0$, the theorem will be re-derived using a separation into two estimates, one for the stochastic terms and one with only measurement-noise for the unknown initial value.

3. SEPARATION INTO TWO ESTIMATES.

The formula from Section 2:

$$\hat{x}(t|t) = \phi(t, t_0) N^T \hat{\xi} + \hat{\beta}(t)$$

shows that $\hat{x}$ can be written as a sum of two estimates. It seems reasonable to separate the state into a stochastic term and a deterministic ξ -term like in (2.5):

$$x = x_1 + x_2 \quad (3.1)$$

$$\begin{cases} x_1(t+1) = \phi x_1(t) + v(t) & x_1(t_0) = x_0^S \\ y_1(t) = \theta x_1(t) + e(t) \end{cases} \quad (3.2)$$

$$\begin{cases} x_2(t+1) = \phi x_2(t) & x_2(t_0) = x_0^N = N^T \xi \\ y_2(t) = \theta x_2(t) \end{cases} \quad (3.3)$$

The Kalman filter for (3.2) would be

$$\hat{x}_1(t+1|t) = \phi \hat{x}_1(t|t-1) + \phi K_{\Pi}(t) [y_1(t) - \theta \hat{x}_1(t|t-1)]$$

$$\hat{x}_1(t_0|t_0-1) = 0$$

$$K_{\Pi}(t) = \Pi(t) \theta^T (\theta \Pi(t) \theta^T + R_2)^{-1} \quad (3.4)$$

$$\Pi(t+1) = \phi [\Pi(t) - K_{\Pi} \theta \Pi(t)] \phi^T + R_1 \quad \Pi(t_0) = R_0^S \quad (3.5)$$

or in operator form

$$\hat{x}_1(t+1|t) = R_{x_1(t+1)Y_{1t}} R_{Y_{1t}}^{-1} Y_{1t}$$

Since y_1 is not available for measurement it is interesting to define $\hat{x}_\Pi$ as the same linear operator applied to y instead:

$$\hat{x}_\Pi(t+1|t) = R_{x_1(t+1)Y_1t} R_{Y_1t}^{-1} Y_t$$

or

$$\hat{x}_\Pi(t+1|t) = \phi \hat{x}_\Pi(t|t-1) + \phi K_\Pi(t) [y(t) - \theta \hat{x}_\Pi(t|t-1)]$$

$$\hat{x}_\Pi(t_0|t_0-1) = 0 \quad (3.6)$$

Assume for a moment that ξ is a stochastic variable independent of v , e and x_0^S . Then by the projection theorem

$$\hat{x}(t+1|t) = R_{x(t+1)Y_t} R_{Y_t}^{-1} Y_t$$

which will be expressed in $\hat{x}_\Pi(t+1|t)$. Drop the time indices:

$$\begin{aligned} \hat{x} &= (R_{x_1Y_1} + R_{x_2Y_2}) R_Y^{-1} Y = R_{x_1Y_1} [R_{Y_1}^{-1} - R_{Y_1}^{-1} R_{Y_2} R_{Y_2}^{-1}] Y + R_{x_2Y_2} R_Y^{-1} Y = \\ &= \hat{x}_\Pi + (R_{x_2Y_2} - R_{x_1Y_1} R_{Y_1}^{-1} R_{Y_2}) R_Y^{-1} Y \end{aligned}$$

by linearity. Since (3.2) and (3.3) are independent

$$R_{x_2Y_2} = R_{x_2Y} \text{ and } R_{Y_2} = R_{Y_2Y} \text{ and since}$$

$$\hat{x}_\Pi - \hat{x}_1 = R_{x_1Y_1} R_{Y_1}^{-1} Y_2$$

it follows that

$$R_{x_2Y} - R_{x_1Y_1} R_{Y_1}^{-1} R_{Y_2Y} = R_{ZY}$$

if z is defined as

$$z(t) = x_2(t) - \hat{x}_\Pi(t|t-1) + \hat{x}_1(t|t-1) \quad (3.7)$$

The projection theorem gives $\hat{z} = R_{zY} R_Y^{-1} Y$ so that

$$\hat{x} = \hat{x}_\Pi + \hat{z} = \hat{x}_\Pi + R_{zY} R_Y^{-1} Y \quad (3.8)$$

Introduce also the equivalent measurement n

$$n = y - \theta \hat{x}_\Pi \quad (3.9)$$

It can be shown that z and n satisfy a simple dynamic system:

Theorem 2: z and n defined by (3.7) and (3.9) satisfy the system

$$\begin{cases} z(t+1) = \phi(I - K_\Pi \theta) z(t) & z(t_0) = x_0^N = N^T \xi \\ n(t) = \theta z(t) + \varepsilon(t) \end{cases} \quad (3.10)$$

where ε is white noise with covariance $\{\theta \Pi(t) \theta^T + R_2\}$.
 K_Π and Π are defined by (3.4) and (3.5).

Proof:

$$\begin{aligned} z(t+1) &= x_2(t+1) - \hat{x}_\Pi(t+1|t) + \hat{x}_1(t+1|t) = \\ &= \phi[x_2(t) - \hat{x}_\Pi(t|t-1) + \hat{x}_1(t|t-1)] - \\ &\quad - \phi K_\Pi(t)[y(t) - \theta \hat{x}_\Pi(t|t-1) - y_1(t) + \theta \hat{x}_1(t|t-1)] = \\ &= \phi z(t) - \phi K_\Pi(t)[y_2(t) + \theta z(t) - \theta x_2(t)] = \\ &= [\phi - \phi K_\Pi(t) \theta] z(t) \end{aligned}$$

$$z(t_0) = x_2(t_0) - \hat{x}_\Pi(t_0|t_0-1) + \hat{x}_1(t_0|t_0-1) = x_2(t_0)$$

$$\eta(t) = y(t) - \theta \hat{x}_{\Pi}(t|t-1) = y_1(t) - \theta \hat{x}_1(t|t-1) + \theta z(t)$$

The innovations $\varepsilon(t) = y_1(t) - \theta \hat{x}_1(t|t-1)$ are white with covariance $R_2 + \theta \Pi(t) \theta^T$.

□

If the covariance of ξ goes to infinity, $\hat{x}_{\Pi}$ is not affected. The original problem with unknown ξ could thus be resumed. $\hat{x}$ is the minimal variance unbiased estimate of x , if $\hat{z}$ is so of z . Theorem 2 is not influenced by the different interpretations of ξ .

Define the estimation errors $\tilde{x}$, $\tilde{x}_1$ and $\tilde{z}$. Then

$$\tilde{x} = x - \hat{x} = x_1 + x_2 - \hat{x}_{\Pi} - \hat{z} = \tilde{x}_1 + \tilde{z} \quad (3.11)$$

Since $\tilde{x}_1$ and $\tilde{z}$ are uncorrelated, define Σ as the covariance of $\tilde{z}$ so that

$$\begin{aligned} \text{Cov } \tilde{x}(t|t-1) &= P(t) = \text{cov } \tilde{x}_1(t|t-1) + \text{cov } \tilde{z}(t|t-1) = \\ &= \Pi(t) + \Sigma(t) \end{aligned} \quad (3.12)$$

In order to get recursive formulas for the estimate $\hat{x}$, formulas must be obtained for $\hat{z}$. The system (3.10) contains only measurement noise. z is deterministic, but ξ is unknown. Such systems will be discussed in the next section.

4. MEASUREMENT NOISE ONLY.

The estimation problem is now brought back to the special case, $v = 0$, $x_0^S = 0$.

In Section 2 the Gauss Markov Theorem was used to express $\hat{\xi}$, given Y_t . The relations (2.2) and (2.3) are now simple since

$$y(t) = \theta\phi(t, t_0)N^T\xi + e(t)$$

and

$$W_t^T Q_t^{-1} W_t = \sum_{s=t_0}^t N\phi^T(s, t_0)\theta^T R_2^{-1}\theta\phi(s, t_0)N^T = NM_{t+1}N^T \quad (4.1)$$

$$W_t^T Q_t^{-1} Y_t = \sum_{s=t_0}^t N\phi^T(s, t_0)\theta^T R_2^{-1}y(s) = N\lambda_{t+1} \quad (4.2)$$

by obvious definitions of M and λ , so that

$$\hat{\xi}(t) = (NM_{t+1}N^T)^{-1}N\lambda_{t+1}$$

Minimal bias estimates: It is possible to get an unbiased estimate $\hat{\xi} = FY$ only if (4.1) is invertible. If not it is only possible to estimate some linear combinations of ξ without bias. Those components of ξ that lie in the null-space of W cannot be estimated without bias. There is, however, freedom left to decide, without knowledge of ξ , in what complementary subspace the estimate should be unbiased. Any such estimate is called a minimum bias estimate. If the rows of W are linear dependent, the freedom should be used to minimize the variance of the estimate. A minimal variance minimal bias linear estimate is thus obtained by the orthogonal pseudo inverse, see also [8].

Thus

$$\hat{\xi}_m = (NMN^T)^+ N \lambda_{t+1} \quad (4.3)$$

for which the components of ξ in the range space of W^T will be estimated without bias.

By (2.4) a good state estimate is

$$\hat{x}_m(t|t) = \phi(t, t_0) N^T \hat{\xi}_m(t)$$

For the degenerate case when Q_t is not invertible similar formulas exist [1] but they will not be considered here. The proofs and the interpretation in the sequel would be more complicated although the final result may be similar.

The estimates can also be obtained by some minimization of the mean square error V in some norm $\| \cdot \|_q$ induced by a quadratic form $x^T q x$

$$V = E \| \xi - FY \|_q^2 = E \| \xi - FW\xi - Fe \|_q^2 = \| \xi - FW\xi \|_q^2 + E \| Fe \|_q^2$$

which, of course, cannot be done directly if nothing is known about ξ . The min max estimate is

$$\hat{\xi}_q = q^{-1} NMN^T (NMN^T q^{-1} NMN^T)^+ N \lambda \quad (4.3a)$$

V is minimized for the worst ξ . The bias term is first minimized, then the variance term.

Note that $\hat{\xi}_q = \hat{\xi}_m$ if $q = I$. If ξ is observable $\hat{\xi}_q = \hat{\xi}_m$ for all q .

It is easily verified that $\hat{\xi}_m$ is the minimal variance unbiased estimate of $NMN^T(NMN^T)^+\xi$, and $\tilde{\xi}_m$ defined by

$$\tilde{\xi}_m = NMN^T(NMN^T)^+\xi - \hat{\xi}_m \quad (4.4)$$

has covariance $(NMN^T)^+$. The bias of $\hat{\xi}_m$ as an estimate of ξ is

$$E(\xi - \hat{\xi}_m) = [I - NMN^T(NMN^T)^+]\xi \quad (4.5)$$

Introduce also $\tilde{x}_m$

$$\tilde{x}_m(t|t) = \phi(t, t_0)N^T\tilde{\xi}_m(t) \quad (4.6)$$

with covariance $\phi(t, t_0)N^T(NMN^T)^+N\phi^T(t, t_0)$.

Example:

$$y = W\xi + e = \begin{bmatrix} 1 & 1 \end{bmatrix}\xi + e \quad Ee = 0 \quad Ee^2 = 1$$

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad M^+ = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \hat{\xi}_m = M^+W^Ty = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} y$$

$$\tilde{\xi}_m = MM^+\xi - \hat{\xi}_m = -\frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} e \quad \text{cov } \tilde{\xi}_m = M^+ = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\tilde{\xi}_{mb} = (I - MM^+)\xi = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \xi$$

$$\min_F E \| \xi - FY \|_Q^2 = \min_F \left\{ \| \xi - FW\xi \|_Q^2 + E \| Fe \|_Q^2 \right\} =$$

$$= \| \xi_0 \|_Q^2 + \min_F \left\{ \| \xi_1 - FW\xi_1 \|_Q^2 + \| F \|_Q^2 \right\}$$

where $W\xi_0 = 0$, $\langle \xi_1, \xi_0 \rangle_q = \xi_1^T q \xi_0 = 0$, $q\xi_1 \in R(W^T)$,

$$\xi = \xi_0 + \xi_1.$$

E.g. $q = I$:

$$F_m = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \xi_1 - FW\xi_1 = 0 \quad \|\xi_0\|^2 = \frac{1}{2} \xi^T \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \xi$$

$$\|F\| = 1/2$$

E.g. $q = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$:

$$F_q = \frac{2}{3} \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \quad \hat{\xi}_q = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} y \quad \hat{\xi}_{q^*} = -\frac{1}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} e$$

$$\hat{\xi}_{q^*} = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \xi \quad \|\xi_0\|_q^2 = \frac{1}{3} \xi^T \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \xi$$

$$\|F\|_q = 5/9$$

Recursive equations: The pseudo inverse $(NMN^T)^+$ can be evaluated recursively using formulas derived by Cline [3], and both $\hat{\xi}_m$ and $\hat{x}_m$ satisfy difference equations like the Kalman filters.

Theorem 3: The minimal variance linear estimate of the state x of (2.1) obtained from the minimal variance, minimal bias, linear estimate of ξ

$$\hat{x}_m(t|t) = \phi(t, t_0) N^T \hat{\xi}(t)$$

satisfies the recursion:

$$\begin{cases} \hat{x}_m(t|t) = \hat{x}_m(t|t-1) + K(t)[y(t) - \theta \hat{x}_m(t|t-1)] \\ \hat{x}_m(t+1|t) = \phi \hat{x}_m(t|t) \end{cases} \quad (4.7)$$

where

$$\begin{aligned} K = \Lambda \theta^T (\theta \Lambda \theta^T)^+ & \left\{ I - (\theta P_m \theta^T + R_2) [A (\theta P_m \theta^T + R_2) A]^+ \right\} + \\ & + P_m \theta^T [A (\theta P_m \theta^T + R_2) A]^+ \end{aligned} \quad (4.8)$$

$$A = I - (\theta \Lambda \theta^T) (\theta \Lambda \theta^T)^+ \quad (4.9)$$

An alternative expression for K is

$$\begin{aligned} K = (\Lambda \theta^T (\theta P_m \theta^T + R_2)^{-1} \theta \Lambda)^+ & \Lambda \theta^T (\theta P_m \theta^T + R_2)^{-1} + \\ & + P_m \theta^T [A (\theta P_m \theta^T + R_2) A]^+ \end{aligned} \quad (4.8a)$$

$P_m(t)$, the covariance of

$$\hat{x}_m(t|t-1) = \phi(t, t_0) N^T [N M_t N^T (N M_t N^T)^+ \xi - \hat{\xi}_m(t-1)]$$

satisfies

$$P_m(t+1) = \phi [(I - K\theta) P_m (I - K\theta)^T + K R_2 K^T] \phi^T, \quad P_m(t_0) = 0 \quad (4.10)$$

The matrix Λ defined by

$$\Lambda(t) = \phi(t, t_0) N^T [I - N M_t N^T (N M_t N^T)^+] N \phi^T(t, t_0) \quad (4.11)$$

satisfies

$$\Lambda(t+1) = \phi [\Lambda - \Lambda \theta^T (\theta \Lambda \theta^T)^+ \theta \Lambda] \phi^T, \quad \Lambda(t_0) = N^T N \quad (4.12)$$

Proof: Introduce some notation:

$$r^T r = R_2^{-1}, \quad c_t = N_\phi^T(t, t_0) \theta^T r^T$$

so that

$$N \lambda_t = \sum_{s=t_0}^{t-1} c_s r y(s)$$

$$\bar{M}_t = E X_{t-1}^m = \sum_{s=t_0}^{t-1} c_s c_s^T$$

$$\hat{\xi}_m(t-1) = \bar{M}_t^+ N \lambda_t$$

$$\hat{\xi}_m(t) = [\bar{M}_t + c_t c_t^T]^+ [N \lambda_t + c_t r y(t)]$$

Now use the pseudo inverse formula and delete indices t :

$$(\bar{M} + c c^T)^+ = \{ (D c c^T D)^+ + [I - (c^T D)^+ c^T] \cdot$$

$$\cdot [\bar{M}^+ - \bar{M}^+ c G B G c^T \bar{M}^+] [I - c (D c)^+] \}$$

$$\text{where } D = I - \bar{M}^+ \bar{M}, \quad G = I - (D c)^+ D c, \quad B = [I + G c^T \bar{M}^+ c G]^{-1}$$

Note that D is the projection on $R(\bar{M})^\perp$ so that $D_t c_s = 0$, $s < t$ giving $D_t N \lambda_t = 0$. Thus

$$\begin{aligned} \hat{\xi}_m(t) &= [I - (c^T D)^+ c^T] [\bar{M}^+ - \bar{M}^+ c G B G c^T \bar{M}^+] N \lambda + \\ &\quad + \{ (D c c^T D)^+ + [I - (c^T D)^+ c^T] [\bar{M}^+ - \bar{M}^+ c G B G c^T \bar{M}^+] \cdot \\ &\quad \cdot [c - c (D c)^+ c] \} r y(t) \end{aligned}$$

but $c - c (D c)^+ c = c - c (D c)^+ D c = c G$ and $(D c c^T D)^+ c = (D c c^T D)^+ D c = (c^T D)^+ = D c (c^T D c)^+$ and $[\bar{M}^+ - \bar{M}^+ c G B G c^T \bar{M}^+] c G = \bar{M}^+ c G [I - B G c^T \bar{M}^+ c G] G = \bar{M}^+ c G B G$. Hence

$$\begin{aligned} \hat{\xi}_m(t) &= \bar{M}^+ N \lambda + \{ D c (c^T D c)^+ + [I - D c (c^T D c)^+ c^T] \bar{M}^+ c G B G \} \cdot \\ &\quad \cdot [r y - c^T \bar{M}^+ N \lambda] \end{aligned}$$

Introduce K'

$$\begin{aligned} K' &= \{ D c (c^T D c)^+ [I - c^T \bar{M}^+ c G B G] + \bar{M}^+ c G B G \} r = \\ &= \{ D c (c^T D c)^+ [I - (I + c^T \bar{M}^+ c) G B G] + \bar{M}^+ c G B G \} r \end{aligned}$$

$$\hat{\xi}_m(t) = \hat{\xi}_m(t-1) + K'(t) [y(t) - c_0^T(t, t_0) N^T \hat{\xi}_m(t-1)] \quad (4.13)$$

The pseudo inverse algebra collected in Appendix takes care of the correlation $R_2 = r^{-1} r^{-T}$ in the expression of K' , so that from (A5)

$$\begin{aligned} K' &= D c r^{-T} (r^{-1} c^T D c r^{-T})^+ \{ I - (R_2 + r^{-1} c^T \bar{M}^+ c r^{-T}) \cdot \\ &\quad \cdot [A (R_2 + r^{-1} c^T \bar{M}^+ c r^{-T}) A]^+ \} + \bar{M}^+ c r^{-T} [A (R_2 + r^{-1} c^T \bar{M}^+ c r^{-T}) A] \\ &\quad \cdot [A (R_2 + r^{-1} c^T \bar{M}^+ c r^{-T}) A]^+ \end{aligned} \quad (4.14)$$

where $A = I - (r^{-1}c^T D c^{-T})(r^{-1}c^T D c^{-T})^+$

Now (2.4) applied to (4.13) gives (4.7) with $K(t) = \phi(t, t_0) N^T K'(t)$.

Λ defined by (4.11) can be written $\Lambda(t) = \phi(t, t_0) D_t N \phi^T(t, t_0)$ and the covariance $P_m(t) = \phi(t, t_0) R^{-1} \bar{M}_t^T \bar{M}_t \phi^T(t, t_0)$ by (4.4). This directly proves (4.8) and (4.9).

It remains to prove (4.8a) and the recursions (4.10) and (4.12). Since $R_2 > 0$ eq. (4.8a) is a rather direct consequence of Lemma 3 in Appendix. Now regard

$$\begin{aligned} I - D_{t+1} &= \bar{M}_{t+1}^+ \bar{M}_{t+1} = [I - (c^T D)^+ c^T] [\bar{M}_t^+ - \bar{M}_t^+ c G B G c^T \bar{M}_t^+] \bar{M}_t + \\ &\quad + (c^T D)^+ c^T + [I - (c^T D)^+ c^T] \bar{M}_t^+ c G B G c^T = \\ &= [I - (c^T D)^+ c^T] [I - D] + (c^T D)^+ c^T - \\ &\quad - [I - (c^T D)^+ c^T] \bar{M}_t^+ c G B G c^T (I - D) - \bar{M}_t^+ c G B G c^T = \\ &= I - D + (c^T D)^+ c^T D + [I - (c^T D)^+ c^T] \bar{M}_t^+ c G B G c^T D \end{aligned}$$

But $G c^T D = 0$ so $D_{t+1} = D - D c (c^T D c)^+ c^T D$. The algebra in Appendix gives

$$\begin{aligned} D c (c^T D c)^+ c^T D &= D c (I - G) r^{-T} (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G) c^T D = \\ &= D c r^{-T} (r^{-1} c^T D c r^{-T})^+ r^{-1} c^T D \end{aligned}$$

and (4.12) follows immediately.

The recursion for P_m is derived from recursions for $\hat{x}_m$ and $\hat{\xi}_m$.

$$\begin{aligned} \hat{x}_m(t) &= \bar{M}_{t+1}^+ M_{t+1} \xi - \hat{\xi}_m(t) = \bar{M}_t^+ \bar{M}_t \xi + D c (c^T D c)^+ c^T D \xi - \\ &\quad - \hat{\xi}_m(t-1) - K'(t) [y(t) - \theta \phi(t, t_0) \hat{\xi}_m(t-1)] = \\ &= \hat{\xi}_m(t-1) - K'(t) [\theta \hat{x}_m(t|t-1) + e(t) + \theta \phi(t, t_0) N^T D \xi] + \\ &\quad + D c (c^T D c)^+ c^T D \xi \end{aligned}$$

But since $G c^T D = 0$

$$\begin{aligned} K' \theta \phi(t, t_0) N^T D &= K' r^{-1} c^T D = \{D c (c^T D c)^+ [I - c^T \bar{M}_t^+ c G B G] + \\ &\quad + \bar{M}_t^+ c G B G\} c^T D = D c (c^T D c)^+ c^T D \end{aligned}$$

and thus

$$\hat{x}_m(t+1|t) = \phi [I - K(t) \theta] \hat{x}_m(t|t-1) - \phi K(t) e(t) \quad (4.15)$$

$$P_m(t+1) = \phi^T \{ [I - K\phi] P_m [I - K\phi]^T + K R_2 K^T \} \phi^T$$

which concludes the proof of the theorem. $\square$

Comments: In the considered special case Theorem 1 and Theorem 3 give the same estimate. Theorem 3 also gives the interpretation of P_m as a covariance matrix and shows that $\hat{x}_m$ is obtained from the minimal bias estimate $\hat{\xi}_m$ of the unknown initial value. It is very natural to assume that $N^T N$ is a projection, which means that the unknown parts of the initial value are "equally unknown". For instance by direct verification of the pseudo inverse conditions it then follows that

$$N^T (NMN^T)^+ N = (N^T NMN^T N)^+$$

and

$$N^T NMN^T N (N^T NMN^T N)^+ = N^T N M M^+ N^T N$$

so that

$$N^T D N = N^T [I - NMN^T (NMN^T)^+] N = N^T N [I - M M^+] N^T N$$

$$x(t) = \phi(t, t_0) N^T N [I - M_t M_t^+] N^T N \phi^T(t, t_0) \quad (4.16)$$

Thus $N^T D N$ is the projection on the unobservable part of the unknown initial value. A is the projection transformed to $x(t)$.

In the next section the full problem with also x_0^S and v will be treated, and the interpretation of the estimate and the matrices P_m and A will be similar.

5. COMBINATION OF TWO FILTERS.

Now return to the original problem (2.1) and to the separation made in Section 3. It was there shown that

$$\hat{x} = \hat{x}_{\Pi} + \hat{z}$$

where $\hat{x}_{\Pi}$ was the stochastic term (3.6) and $\hat{z}$ the bias term. According to Theorem 2 z was the state of a system with only white measurement noise. Such systems were treated in Section 4, and a good minimal variance estimate of z is obtained from the minimal variance, minimal bias linear estimate of ξ .

$$\hat{z}_m(t|t) = \left\{ \prod_{s=t_0}^{t-1} (\phi - \phi K_{\Pi}(s)\theta) \right\} \hat{\xi}_m(t) = \psi(t, t_0) \hat{\xi}_m(t) \quad (5.1)$$

which satisfies

$$\begin{aligned} \hat{z}_m(t|t) &= \hat{z}_m(t|t-1) + K_z(t)[y(t) - \theta \hat{x}_{\Pi}(t|t-1) - \\ &\quad - \theta \hat{z}_m(t|t-1)] \end{aligned} \quad (5.2)$$

$$\hat{z}_m(t+1|t) = [\phi - \phi K_{\Pi}(t)\theta] \hat{z}_m(t|t)$$

$$\begin{aligned} K_z &= \Lambda \theta^T (\theta \Lambda \theta^T)^+ \left\{ I - (\theta \Sigma_m \theta^T + \theta \Pi \theta^T + R_2) [A(\theta \Sigma_m \theta^T + \theta \Pi \theta^T + R_2) A]^+ \right\} + \\ &\quad + \Sigma_m \theta^T [A(\theta \Sigma_m \theta^T + \theta \Pi \theta^T + R_2) A]^+ \end{aligned} \quad (5.3)$$

$$A = I - \theta \Lambda \theta^T (\theta \Lambda \theta^T)^+$$

$$\Lambda(t+1) = \phi(I - K_{\Pi}\theta) [\Lambda - \Lambda \theta^T (\theta \Lambda \theta^T)^+ \theta \Lambda] (I - K_{\Pi}\theta)^T \phi^T$$

$$\Lambda(t_0) = N^T N \quad (5.4)$$

$$\begin{cases} \Sigma_m(t+1) = \phi(I-K_{\Pi}\theta) \left\{ (I-K_Z\theta) \Sigma_m(I-K_Z\theta)^T + \right. \\ \left. + K_Z(R_2+\theta\Pi\theta^T)K_Z^T \right\} (I-K_{\Pi}\theta)^T \phi^T \\ \Sigma_m(t_0) = 0 \end{cases} \quad (5.5)$$

where Σ_m is the covariance of $\hat{z}_m(t|t-1)$

$$\hat{z}_m(t|t-1) = \psi(t, t_0) N^T [N M_t N^T (N M_t N^T)^+ \xi - \hat{\xi}(t-1)] \quad (5.6)$$

with

$$M_t = \sum_{s=t_0}^{t-1} \psi^T(s, t_0) \theta^T (\theta \Pi(s) \theta^T + R_2)^{-1} \theta \psi(s, t_0) \quad (5.7)$$

It can now be proven that (3.8) gives the same estimate as Theorem 1:

Theorem 4: The minimal variance linear estimate of the state x of (2.1) obtained from the minimal variance, minimal bias, linear estimate of the unknown initial state ξ by

$$\hat{x}_m(t|t) = \hat{x}_{\Pi}(t|t) + \hat{z}_m(t|t) \quad (5.8)$$

with $\hat{x}_{\Pi}$ defined by (3.6) and $\hat{z}_m$ by (5.1) satisfies the recursion

$$\hat{x}_m(t|t) = \hat{x}_m(t|t-1) + K(t)[y(t) - \theta \hat{x}_m(t|t-1)] \quad (5.9)$$

$$\hat{x}_m(t+1|t) = \phi \hat{x}_m(t|t)$$

where

$$\begin{aligned} K = \Lambda \theta^T (\theta \Lambda \theta^T)^+ \left\{ I - (\theta P_m \theta^T + R_2) [A(\theta P_m \theta^T + R_2) A]^+ \right\} + \\ + P_m \theta^T [A(\theta P_m \theta^T + R_2) A]^+ \end{aligned} \quad (5.10)$$

$$A = I - (\theta \Lambda \theta^T)(\theta \Lambda \theta^T)^+ \quad (5.11)$$

An alternative expression for K is

$$K = [\Lambda \theta^T (\theta P_m \theta^T + R_2)^{-1} \theta \Lambda]^+ \Lambda \theta^T (\theta P_m \theta^T + R_2)^{-1} + \\ + P_m \theta^T [A (\theta P_m \theta^T + R_2) A]^+ \quad (5.10a)$$

$P_m(t)$, the covariance of $\hat{x}_m(t|t-1)$

$$\hat{x}_m(t|t-1) = \hat{x}_1(t|t-1) + \hat{z}_m(t|t-1) \quad (5.12)$$

with $\hat{z}_m$ from (5.6) satisfies

$$P_m(t+1) = R_1 + \phi [(I - K\theta) P_m (I - K\theta)^T + K R_2 K^T] \phi^T \\ P_m(t_0) = R_0^S \quad (5.13)$$

and Λ , the transformed projection on the unobservable part of the unknown initial value satisfies

$$\Lambda(t+1) = \phi [\Lambda - \Lambda \theta^T (\theta \Lambda \theta^T)^+ \theta \Lambda] \phi^T \quad \Lambda(t_0) = N^T N \quad (5.14)$$

Proof:

$$\hat{x}_m(t|t) = \hat{x}_\Pi(t|t) + \hat{z}_m(t|t) = \hat{x}_\Pi(t|t-1) + \hat{z}_m(t|t-1) + \\ + K_\Pi [y - \theta \hat{x}_\Pi] + (I - K_\Pi \theta) K_Z (y - \theta \hat{x}_\Pi - \theta \hat{z}_m) = \\ = \hat{x}(t|t-1) + K(t) [y(t) - \theta \hat{x}(t|t-1)]$$

with $K = K_\Pi + (I - K_\Pi \theta) K_Z$. Note that $(I - K\theta) = (I - K_\Pi \theta)(I - K_Z \theta)$.

The covariance P_m of $\hat{x}_m$ is $P_m = \Pi + \Sigma_m$ so that by (5.5) and (3.5)

$$P_m(t+1) = R_1 + \phi \{ (I - K_\Pi \theta) \Pi + (I - K_\Pi \theta)(I - K_Z \theta) \Sigma_m (I - K_Z \theta)^T \cdot \\ \cdot (I - K_\Pi \theta)^T + (I - K_\Pi \theta) K_Z (R_2 + \theta \Pi \theta^T) K_Z^T (I - K_\Pi \theta)^T \} \phi = \\ = R_1 + \phi \{ (I - K\theta) P_m (I - K\theta)^T + K R_2 K^T \} \phi$$

Since $(I-K\theta)\Pi - (I-K\theta)\Pi(I-K\theta)^T + (K-K_\Pi)(R_2+\theta\Pi\theta^T)(K-K_\Pi)^T - KR_2K^T = 0$ which follows from:

$$(I-K\theta)\Pi(I-K\theta)^T = \Pi - K\theta\Pi - \Pi\theta^TK^T + K\theta\Pi\theta^TK^T$$

$$(K-K_\Pi)(R_2+\theta\Pi\theta^T)(K-K_\Pi)^T = KR_2K^T + K\theta\Pi\theta^TK^T - (K-K_\Pi) \cdot$$

$$\begin{aligned} & \cdot (R_2+\theta\Pi\theta^T)K_\Pi^T - K_\Pi(R_2+\theta\Pi\theta^T)K^T = \\ & = KR_2K^T + K\theta\Pi\theta^TK^T - (K-K_\Pi)\theta\Pi - \Pi\theta^TK^T \end{aligned}$$

Λ defined by (5.4) also fulfils (5.14), since

$$\theta(\Lambda - \Lambda\theta(\theta\Lambda\theta^T)^+\theta\Lambda) = 0.$$

It remains to prove that $K = K_\Pi + (I-K_\Pi\theta)K_Z$ can be evaluated by (5.10). Then (5.10a) follows like in Theorem 3. Denote $R_2 + \theta\Pi\theta^T + \theta\Sigma_m\theta^T$ by R :

$$\begin{aligned} K &= K_Z + K_\Pi(I-\theta K_Z) = \Sigma_m\theta^T[ARA]^+ + \Lambda\theta^T(\theta\Lambda\theta^T)^+\{I - R[ARA]^+\} + \\ &+ K_\Pi\{I - \theta\Sigma_m\theta^T[ARA]^+\} - K_\Pi\theta\Lambda\theta^T(\theta\Lambda\theta^T)^+[I - R(ARA)^+] = \\ &= (\Pi+\Sigma_m)\theta^T[ARA]^+ + \Lambda\theta^T(\theta\Lambda\theta^T)^+\{I - R[ARA]^+\} + \\ &+ K_\Pi\{I - [(\theta\Pi\theta^T+R_2) + \theta\Sigma_m\theta^T][ARA]^+\} + K_\Pi[I - (\theta\Lambda\theta^T) \cdot \\ &\cdot (\theta\Lambda\theta^T)^+]\{I - R[ARA]^+\} - K_\Pi\{I - R(ARA)^+\} = \\ &= P_m\theta^T[ARA]^+ + \Lambda\theta^T(\theta\Lambda\theta^T)^+\{I - R[ARA]^+\} \end{aligned}$$

$$\text{since } A(I - R(ARA)^+) = 0$$

□

Comments: In order to get a correct interpretation of Λ in (5.14) and (5.4) it must be noticed that the unobservable subspace is the same for the z-system of Theorem 2 and the original system (2.1), so both $I-MM^+$ and Λ are the same! Thus if N^TN is a projection,

$$\Lambda(t) = \phi(t, t_0)N^TN(I-M_t^+M_t)N^TN\phi^T(t, t_0) \quad (5.15)$$

$$M_t = \sum_{s=t_0}^{t-1} \phi^T(s, t_0)\theta^T\theta\phi(s, t_0) \quad (5.16)$$

In the same way with $\hat{z}_m(t|t-1) = \psi(t, t_0) N^T \hat{\xi}_m(t-1)$ and with $\tilde{x}_m = \tilde{x}_1 + \tilde{z}_m$ the bias term can be written

$$\begin{aligned} E(x - \hat{x}) &= \tilde{x}_{mb}(t|t-1) = x(t) - \hat{x}_m(t|t-1) - \tilde{x}_m(t|t-1) = \\ &= z - \hat{z}_m - \tilde{z}_m = \psi(t, t_0) N^T N (I - M_t M_t^+) N^T \xi = \\ &= \phi(t, t_0) N^T N (I - M_t M_t^+) N^T \xi \end{aligned} \quad (5.17)$$

so that Λ spans the bias. P_m is the covariance of $x - \hat{x}_m$ since

$$P_m = E(x - \hat{x} - \tilde{x}_{mb})^2$$

6. LINEAR STOCHASTIC CONTROL.

An application of the above filter results is the control of a linear stochastic system for which a quadratic loss function:

$$J = E \sum_{t=t_0}^{N-1} \left\{ x^T(t) Q_1 x(t) + u^T(t) Q_2 u(t) \right\} + x^T(N) Q_0 x(N) \quad (6.1)$$

should be minimized with respect to $u(t_0), \dots, u(N)$ subject to the constraint

$$\begin{cases} x(t+1) = \phi x(t) + \Gamma u(t) + v(t), & x(t_0) = x_0 = x_0^S + N^T \xi \\ y(t) = \theta x(t) + R(t) \end{cases} \quad (6.2)$$

with v , e , x_0 defined in Section 2. The expectation in (6.1) is taken with respect to the introduced statistics v , e and x_0^S . The choice of $u(t)$ is restricted to linear maps of Y_{t-1} .

Rewrite J as in [2] using S and L defined by

$$L(t) = (Q_2 + \Gamma^T S(t+1) \Gamma)^{-1} \Gamma^T S(t+1) \phi \quad (6.3)$$

$$S(t) = \phi^T S(t+1) \phi - \phi^T S(t+1) \Gamma L(t) + Q_1, \quad S(N) = Q_0 \quad (6.4)$$

$$J = E x_0^T S(t_0) x_0 + E \sum_{t=t_0}^{N-1} \left\{ v^T(t) S(t+1) v(t) + (u(t) + L(t)x(t))^T \cdot \right.$$

$$\cdot (Q_2 + \Gamma^T S(t+1) \Gamma) (u(t) + L(t)x(t)) =$$

$$= \xi^T N S(t_0) N^T \xi + \text{tr } R_0^S S(t_0) + \sum_{t=t_0}^{N-1} \text{tr } R_1(t) S(t+1) + E \sum_{t=t_0}^{N-1} (u+Lx)^T \cdot$$

$$\cdot (Q_2 + \Gamma^T S \Gamma) (u+Lx) \quad (6.5)$$

Now consider the terms

$$T_t = E \| u(t) + L(t)x(t) \|_{(Q_2 + r^T S(t+1)r)}^2 \quad (6.6)$$

for $u(t)$ being a linear function of Y_{t-1}

$$u(t) = G_t Y_{t-1} \quad (6.7)$$

In order to minimize (6.6) G_t should lie in $R(L)$, since no component orthogonal to $R(L)$, in the scalar product $\langle, \rangle_{Q_2 + r^T S r}$, would decrease T . Thus

$$u = LHY$$

and

$$T = E \| x + HY \|_{L^T(Q_2 + r^T S r)L}^2$$

If the minimum is evaluated for the worst ξ , this is just the problem of minimal variance, minimal bias linear estimates discussed in Section 4. The norm, $\| \cdot \|_q$, in which the error should be measured is here induced by the matrix

$$q = L^T(Q_2 + r^T S r)L \quad (6.8)$$

Rewrite x as

$$x = \hat{x}_q + \tilde{x}_q + \tilde{x}_{qb}$$

where the estimation error is given by the zero mean term $\tilde{x}_q$ and the deterministic bias term is $\tilde{x}_{qb}$. T is then minimized for

$$u = -L\hat{x}_q \quad (6.9)$$

and the minimal T is

$$\begin{aligned} T^0 &= E \| \tilde{x}_q + \tilde{x}_{qb} \|_q^2 = \| \tilde{x}_{qb} \|_q^2 + \text{tr}(q \cdot \text{cov } \tilde{x}_q) = \\ &= \| \tilde{x}_{qb} \|_q^2 + \text{tr} \left\{ P_q L^T (Q_2 + F^T S F) L \right\} \end{aligned}$$

Now $\tilde{x}_q(t|t-1)$ and also $\tilde{x}_{qb}(t|t-1)$ do not change with different $u(s)$, $s < t$, restricted to linear functions of Y_{s-1} , a fact that is fairly easy to show using for instance innovations. Thus the last sum in (6.5) is minimized, starting with the term T_{N-1} , by u from (6.9). The minimal J is thus

$$\begin{aligned} J^0 &= \xi^T N S(t_0) N^T \xi + \text{tr } R_0^S S(t_0) + \sum_{t=t_0}^{N-1} \text{tr } R_1(t) S(t+1) + \\ &+ \sum_{t=t_0}^{N-1} T_t^0 \end{aligned}$$

This concludes the proof of the following theorem.

Theorem 5: The loss function (6.1) for the system (6.2) is minimized for worst possible ξ by the input

$$u(t) = -L(t) \hat{x}_q(t|t-1) \quad (6.10)$$

in the class of linear functions of Y_{t-1} , (6.7). $L(t)$ is defined by (6.3), (6.4), and $\hat{x}_q$ is the minimal variance, minimal bias linear estimate of $x(t)$ with the error measured by the matrix q in (6.8). The minimal loss, which depends on the unknown constant ξ , is

$$\begin{aligned}
J^0 = & \xi^T N S(t_0) N^T \xi + \text{tr } R_0^S S(t_0) + \sum_{t=t_0}^{N-1} \text{tr } R_1(t) S(t+1) + \\
& + \sum_{t=t_0}^{N-1} \text{tr } P_q L^T (Q_2 + r^T G r) L + \sum_{t=t_0}^{N-1} \| \hat{x}_{q_b} \|_q^2 \quad (6.11)
\end{aligned}$$

Comments: The bias terms $\hat{x}_{q_b}$ will be zero as soon as ξ is observable. The restriction (6.7) on u is rather natural since all the random variables have zero mean value, but it can be argued that it implies that possible values of ξ are assumed to be centered around zero. If a bias term is allowed in u , it is no longer meaningful to minimize the bias of $\hat{x}$. The estimate $\hat{x}_m$ of Theorem 1 would be as good as any $\hat{x}_q$, since the estimates will become unbiased at the same time. Since q is time varying there seems to be no hope to obtain simple general recursive formulas for $\hat{x}_q$.

7. CONTINUOUS TIME, DUALITY.

In the continuous time case it is not possible to obtain recursive estimates by letting the covariance increase like in Section 2 or by some pseudo inverse formulas like in Sections 4 and 5. The minimal unbiased estimate will be obtained by duality. Consider the system

$$\begin{cases} \dot{x} = Ax + v \\ y = Cx + e \end{cases} \quad x(t_0) = x_0 = x_0^S + x_0^N \quad (7.1)$$

where v and e are uncorrelated white noise with covariances $R_1\delta(t)$ and $R_2\delta(t)$, $R_2 > 0$. x_0^S is uncorrelated with e and v , has zero mean value and covariance R_0^S . The only thing known about x_0^N is that it is restricted to a subspace spanned by the full column rank rectangular matrix N^T :

$$x_0^N = N^T \xi, \quad \xi \text{ arbitrary}$$

It is a celebrated fact that the filter problem is the dual of an optimization problem [2, 5, 9] and that was used for the proof of the continuous time filter problem in [2]. There is also a well-known duality between observability and controllability, i.e. reconstruction or unbiased estimates and fixed end-point problems [2, 6, 9].

These two dualities will here be shown to combine.

Consider the minimal variance unbiased estimate for the system (7.1). It is convenient first to estimate an arbitrary linear combination of $x(t)$, say $a^T x(t)$. Thus the variance

$$J = E[a^T x(t) - a^T \hat{x}(t)]^2 \quad (7.2)$$

should be minimized for linear functions of Y_+ , say

$$a^T \hat{x}(t) = - \int_0^t u^T(s) y(s) ds \quad (7.3)$$

under the constraint of being unbiased

$$E a^T \hat{x}(t) = a^T E x(t) \quad (7.4)$$

The notation $a^T \hat{x}(t)$ will be justified later. There is usually an $\hat{x}(t)$ such that $a^T \hat{x}(t)$ is the best estimate for $a^T x(t)$.

Theorem 6: The unbiased filter estimate problem for (7.1) is dual to the restricted end-point optimization problem, with the loss function

$$= z^T(t_0) R_0^S z(t_0) + \int_{t_0}^t [z^T(s) R_1 z(s) + u^T(s) R_2 u(s)] ds \quad (7.5)$$

and the constraints

$$\dot{z} = A^T z + C^T u, \quad z(t) = a \quad (7.6)$$

$$N z(t_0) = 0 \quad (7.7)$$

Proof: Use z defined by (7.6), $a^T \hat{x}$ from (7.3) and the adjoint equation (7.1) in the same way as in [2] to reduce the estimation error

$$a^T x(t) - a^T \hat{x}(t) = z^T(t_0) x(t_0) + \int_{t_0}^t z^T(s) v(s) ds + \int_{t_0}^t u^T(s) e(s) ds$$

In order to fulfil (7.4)

$$z^T(t_0)x(t_0) = z^T(t_0)N^T\xi = (Nz(t_0))^T\xi = 0$$

for all ξ , which is equivalent to (7.7). V can thus be expressed as

$$\begin{aligned} V &= E[a^T x(t) - a^T \hat{x}(t)]^2 = \\ &= z^T(t_0)R_0^S z(t_0) + \int_{t_0}^t z^T(s)R_1 z(s) + u^T(s)R_2 u(s) ds \end{aligned}$$

using the covariances of x_0^S , v and e .

□

The solution of the optimization is well-known, cf. [6, 9], but requires controllability of the restricted end-state, i.e. observability of $N^T\xi$.

$$u(s) = -R_2^{-1}C[\Pi(s)z(s) + \psi(s, t_0)N^T(NM(t_0, t)N^T)^{-1}N\psi(t, t_0)a] \quad (7.8)$$

$$\dot{\Pi} = A\Pi + \Pi A^T + R_1 - \Pi C^T R_2^{-1} C \Pi, \quad \Pi(t_0) = R_0^S \quad (7.9)$$

$$\frac{d}{dt} \psi(t, s) = (A - \Pi(t)C^T R_2^{-1} C)\psi(t, s), \quad \psi(s, s) = I \quad (7.10)$$

$$M(t_0, t) = \int_{t_0}^t \psi^T(s, t_0)C^T R_2^{-1} C \psi(s, t_0) ds \quad (7.11)$$

M is an observability Gramian for (7.1). Now solve (7.6), (7.8) for z giving $\hat{x}$ independent of a :

$$\begin{aligned} z(s) &= \left[\psi^T(t, s) - M(s, t)\psi(t, t_0)N^T(NM(t_0, t)N^T)^{-1}N\psi^T(t, t_0) \right] a \\ &= \int_0^t \left[\psi(t, s)\Pi(s) - \psi(t, t_0)N^T(NM(t_0, t)N^T)^{-1} \right. \\ &\quad \cdot N\psi^T(t, t_0)M(s, t)\Pi(s) + \psi(t, t_0)N^T(NM(t_0, t)N^T)^{-1} \\ &\quad \cdot N\psi^T(t, t_0) \left. \right] C^T R_2^{-1} y(s) ds \end{aligned} \quad (7.12)$$

Note that the minimal V is

$$\begin{aligned} a^T [\Pi(t) + \psi(t, t_0) N^T (NM(t_0, t) N^T)^{-1} N \psi^T(t, t_0)] a = \\ = a^T [\Pi(t) + \Sigma(t)] a \end{aligned}$$

defining the error covariance $P(t) = \Pi(t) + \Sigma(t)$.

It is now possible to state the theorem

Theorem 7: The best linear unbiased state estimate for (7.1) is

$$\hat{x}(t|t) = \hat{x}_{\Pi}(t|t) + \psi(t, t_0) N^T (NM(t_0, t) N^T)^{-1} N \lambda(t_0, t) \quad (7.13)$$

with ψ , M and Π from (3.9), (3.10) and (3.8) and $\hat{x}_{\Pi}$ and λ from

$$\begin{aligned} \frac{d}{dt} \hat{x}_{\Pi}(t|t) &= A \hat{x}_{\Pi}(t|t) + \Pi(t) C^T R_2^{-1} [y(t) - C \hat{x}_{\Pi}(t|t)], \\ \hat{x}_{\Pi}(t_0|t_0) &= 0 \end{aligned} \quad (7.14)$$

$$\begin{cases} -\frac{d}{ds} \lambda(s, t) = (A - \Pi(s) C^T R_2^{-1} C) \lambda(s, t) + \\ \quad + C^T R_2^{-1} [y(s) - C \hat{x}_{\Pi}(s|s)] \\ \lambda(t, t) = 0 \end{cases} \quad (7.15)$$

The error covariance is $P(t) = \Pi(t) + \Sigma(t)$.

Proof: Integrate (7.14) and (7.15)

$$\hat{x}_{\Pi}(t|t) = \int_{t_0}^t \psi(t, s) \Pi(s) C^T R_2^{-1} y(s) ds$$

$$\lambda(s,+) = \int_s^t \psi^T(q,s) C^T R_2^{-1} [y(q) - (\hat{x}_\Pi(s|q) - \hat{x}_\Pi(s|s))] dq$$

and combine (7.12) with

$$\begin{aligned} \int_{t_0}^t \psi^T(s,t_0) M(s,t) \Pi(s) C^T R_2^{-1} y(s) ds &= \\ &= \int_{t_0}^t \psi^T(q,t_0) C^T R_2^{-1} C \left[\int_{t_0}^q \psi(q,s) \Pi(s) C^T R_2^{-1} y(s) ds \right] dq \quad \square \end{aligned}$$

Example 7.1: $A = 0$, $R_1 = \sigma^2$, $R_2 = 1$, $x(0) = \xi$ unknown, $C = 1$.

$$\dot{\Pi} = \sigma^2 - \Pi^2, \quad \Pi(0) = 0 \Rightarrow \Pi(t) = \sigma \tanh \sigma t$$

$$\psi(t,s) = \exp \left\{ \int_s^t -\sigma \tanh \sigma q \, dq \right\} = \cosh \sigma s / \cosh \sigma t$$

$$\Sigma(t) = \frac{1}{\cosh^2 \sigma t} / \int_0^t \frac{ds}{\cosh^2 \sigma s} = \sigma / (\sinh \sigma t \cdot \cosh \sigma t) \quad t > 0$$

$$P(t) = \Pi(t) + \Sigma(t) = \sigma \coth \sigma t \quad t > 0$$

$$\hat{x}_\Pi(t|t) = \sigma \int_0^t \frac{\sinh \sigma s}{\cosh \sigma s} y(s) ds$$

$$\hat{x}(t|t) = \hat{x}_\Pi(t|t) + \frac{\sigma}{\sinh \sigma t} \int_0^t \frac{y(s) - \hat{x}_\Pi(s|s)}{\cosh \sigma s} ds =$$

$$= \frac{\sigma}{\sinh \sigma t} \int_0^t \cosh \sigma s y(s) ds \quad t > 0$$

This could be obtained analytically using $P(t)$

$$\begin{aligned}\hat{x}(t|t) &= \int_0^t \exp\left[-\sigma \int_s^t \coth \sigma q dq\right] \sigma \coth \sigma y(s) ds = \\ &= \frac{\sigma}{\sinh \sigma t} \int_0^t \cosh \sigma y(s) ds\end{aligned}$$

but to use a numerical algorithm for the Riccati equation started with a large variance would lead to enormous difficulties.

Example 7.2: $A = -a$, $C = 1$, $R_1 = 0$, $R_2 = 1$

$$r(t) = 0, \quad p(t,s) = p(t,s) = \exp\{-a(t-s)\}$$

$$P(t) = \Sigma(t) = e^{-2at} / \int_0^t e^{-2as} ds = 2a e^{-2at} / (1 - e^{-2at}) \quad t > 0$$

$$\hat{x}(t|t) = 2a e^{-at} \int_0^t e^{-as} y(s) ds / (1 - e^{-2at}) \quad t > 0$$

The initial estimate in both the examples is

$$\lim_{t \rightarrow 0^+} \hat{x}(t|t) = y(0)$$

which has infinite covariance.

Comments: The best unbiased estimate obtained by Theorem 7 is a sum of two estimates, similar to the discrete time case. The last term consists of a transformed smoothing estimate of the initial constant ξ .

It must be emphasized that it is not possible to obtain the estimate from a recursive filter. From (7.13) and the examples it can be seen what happens at the initial point. The estimate may very well exist but with infinite variance. For time invariant systems the whole state becomes observable after an infinitely short time, so even if a pseudo inverse is used instead of P , it is infinite and the gain required in the differential equation at $t = t_0$ is infinite. The discontinuity in $\hat{x}$ and P must be calculated in some other way, preferably from Theorem 7, if a Kalman filter is to be started at $t = t_0^+$.

However, any observability index is very small at t_0^+ and P is very large. It is suggested that the Kalman filter should not be started until a little later for numerical reasons. In fact, $\lambda(t_0, t)$ of (7.15) can be obtained by integration in the forward direction:

$$\frac{d}{dt} \lambda(t_0, t) = \psi^T(t, t_0) C^T R_2^{-1} [y(t) - C \hat{x}_{\Pi}(t|t)],$$

$$\lambda(t_0, t_0) = 0 \quad (7.16)$$

The differential equation is not asymptotically stable, but it does not matter since it is only needed a short time. Thus $\hat{x}$ can be obtained by (7.13) from $\hat{x}_{\Pi}$ and λ , until $P(t)$ is small enough to start the Kalman filter.

It should also be noted, that it is impossible to get any unbiased state estimate with a constant gain Kalman filter in the continuous time case. "Dead-beat" filters must have time varying gain. Compare the dual problem with dead-beat controllers.

6. CONCLUSIONS.

It has been shown how the Kalman filter should be started if part of the initial state is totally unknown. The formulas can be obtained formally by letting the covariance of that part go to infinity. The common way of starting a Kalman filter with a very large covariance is thus almost optimal, but the numerical properties of such a solution are bad.

The optimal discrete time solution uses two "Riccati equations", one to keep track of the bias from the unknown parts of the initial state, and one for the error covariance.

The interpretation of the estimate is provided by a separation into two filters. Before it is possible to obtain an unbiased estimate, the obtained estimate minimizes the mean square error of the unknown initial value in the Euclidian norm. The linear stochastic regulator problem is, however, shown to require an estimate that minimizes the error in another norm. The minimum of the loss function will contain additional terms containing the unknown initial value.

The continuous time case is more complicated. The whole system becomes observable at once. After an initial discontinuity in estimate and covariance a usual Kalman filter could be started. This, however, implies almost infinite Kalman gain and poor numerical properties. It is suggested that the estimate is calculated using separation into two estimates. The "noise term" satisfies a simple filter of "Kalman structure", while the "bias term" should be calculated from a recursively updated smoothing estimate of the unknown initial state. When the error covariance has become reasonable, the usual Kalman filter should be started.

APPENDIX

Although pseudo inverses are conceptually simple some of the necessary algebra tends to obscure the ideas. Some lemmas will be shown here to simplify the proof of Theorem 3. A good general reference is [1].

Lemma 1: Let $\bar{M}$ be a symmetric matrix and c a rectangular matrix, let D and G be the projections $D = I - \bar{M}^+ \bar{M}$, $G = I - (Dc)^+ Dc$, and let r be an invertible matrix, then the projection $A = I - r^{-1} c^T D c r^{-T} (r^{-1} c^T D c r^{-T})^+$ can be obtained by

$$A = r^T G r (r^T G r)^+ \quad (A.1)$$

Proof: If the matrix X spans the subspace orthogonal to that spanned by $c^T D$, i.e. $Gx = x$, then $r^T x \perp r^{-1} c^T D$. $R(A) = R(r^T x) = R(r^T G)$ so that $r^T G r (r^T G r)^+$ is the unique orthogonal projection A . $\square$

Lemma 2: With r , D , G , c , $\bar{M}$ and A as above and with $B = [I + G c^T \bar{M}^+ c G]^{-1}$

$$r^T G B G r = [A (V V^T + R_2) A]^+ \quad (A.2)$$

$$\text{if } r^T c = V_2^{-1} \text{ and } c^{-1} c^T \bar{M}^+ c r^{-T} = V V^T$$

Proof:

$$\begin{aligned} r^T G B G r &= r^T G r [R_2^{-1} + r^T G r V V^T r^T G r]^{-1} r^T G r = \\ &= r^T G r [R_2 - R_2 r^T G r V (I + V^T r^T G r R_2 r^T G r V)^{-1} V^T r^T G r R_2] r^T G r = \\ &= r^T G r - r^T G r V (I + V^T r^T G r V)^{-1} V^T r^T G r \end{aligned}$$

by the inverse lemma and the fact that G is a projection. With $F = r^T G r$, the right hand side of (A.2) gives using Lemma 1 and the pseudo inverse lemma:

$$\begin{aligned} [A(VV^T + R_2)A]^+ &= [F^+ F V V^T (F^+ + F^+ F F^+)]^+ = \\ &= F - F F^+ F V (I + V^T (F^+ F F^+ F V)^{-1} V^T F F^+ F) = \\ &= F - F V (I + V^T F V)^{-1} V^T F \end{aligned}$$

Note the simple form of the pseudo inverse lemma since R_2 invertible and thus $AV \in R(AR_2)$. $\square$

Lemma 3: With c , D , G , A and r as above

$$G = r^{-T} A r^{-1} (r^{-T} A r^{-1})^+ = r^{-T} (A r^{-1} r^{-T} A)^+ r^{-1} \quad (A.3)$$

$$(c^T D c)^+ = (I - G) r^{-T} (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G) \quad (A.4)$$

Proof: The first equality of (A.3) is the inversion of Lemma 1, and the second follows from $A r^{-1} (r^{-T} A r^{-1})^+ = (r^{-T} A)^+ = (A r^{-1} r^{-T} A)^+ A r^{-1}$. Direct verification of the Moore Penrose conditions, $B = A^+$ if AB and BA symmetric and $ABA = A$, $BAB = B$, can be used to show (A.4):

$$\begin{aligned} c^T D c (I - G) r^{-T} (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G) &= \\ &= r (r^{-1} c^T D c r^{-T}) (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G) = \\ &= r (I - A) r^{-1} (I - G) = I - G - r r^T (r^{-T} A r^{-1}) (I - G) = I - G \end{aligned}$$

which is symmetric and $(I - G) c^T D c = c^T D c (I - G) r^{-T} \cdot$
 $\cdot (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G)^2 = (I - G) r^{-T} (r^{-1} c^T D c r^{-T})^+ r^{-1} (I - G).$ $\square$

Lemma 4: The matrix K' defined by (4.13) can be expressed as

$$\begin{aligned} K' &= \left\{ D c (c^T D c)^+ [I - (I + c^T \bar{M}^+ c) G B G] + \bar{M}^+ c G B G \right\} r = \\ &= D c r^{-T} (r^{-1} c^T D c r^{-T})^+ \left\{ I - (R_2 + V V^T) [A (R_2 + V V^T) A]^+ \right\} + \\ &\quad + \bar{M}^+ c r^{-T} [A (R_2 + V V^T) A]^+ \quad (A.5) \end{aligned}$$

to prove (A.5) over.

Proof: Since $DcG = 0$, (A.4) and (A.2) give

$$\begin{aligned} Dc(c^T Dc)^+ [I - (I + c^T \bar{M} + c)GBG]r &= \\ &= Dcr^{-T}(r^{-1}c^T Dcr^{-T})^+(I - r^{-1}Gr)\{I - (R_2 + VV^T)[A(R_2 + VV^T)A]^+\} \end{aligned}$$

From (A.3) $r^{-1}Gr = (AR_2A)^+$, and since R_2 has full rank $AV \in R(AR_2)$ so that

$$\begin{aligned} (AR_2A)^+ \{I - (R_2 + VV^T)[AR_2A + AVV^T A]^+\} &= \\ &= (AR_2A)^+ \{I - A(R_2 + VV^T)A[AR_2A + AVV^T A]^+\} = 0 \end{aligned}$$

which proves (A.5). □

REFERENCES.

- [1] Albert, A.: Regression and the Moore-Penrose Pseudo Inverse, Academic Press, New York, 1972.
- [2] Åström, K.J.: Introduction to Stochastic Control Theory, Academic Press, New York, 1970.
- [3] Cline, R.E.: Representation for the Generalized Inverse of Sums of Matrices. Siam. J. Numer. Anal., Serie B, 2, 99 - 114 (1965).
- [4] Hagander, P.: The Use of Operator Factorization for Linear Control and Estimation, Automatica, 9, 623 - 631 (1973).
- [5] Kalman, R.E.: New Methods in Wiener Filtering Theory, in Proceedings of First Symp. on Eng. Appl. of Random Function Theory and Probability, J.L. Bogdanoff and F. Kozin (eds.), Wiley, New York, 1963.
- [6] Kalman, R.E., Ho, Y.-C, Narendra, K.S.: Controllability of Linear Dynamical Systems, Contributions to Differential Equations, 1, 189-213 (1961).
- [7] Luenberger, D.G.: Optimization by Vector Space Methods, John Wiley, New York, 1969.
- [8] Rao, C.R., Mitra, S.K.: Generalized Inverse of Matrices and its Applications, John Wiley, New York, 1971.
- [9] Sorenson, H.W.: Controllability and Observability of Linear, Stochastic, Time-Discrete Control Systems, in Advances in Control Systems, C.T. Leonides (ed.), Academic Press, New York, 1968.

